ISSN : 2356 - 5195

PREPROCESSING DATA UNTUK SISTEM PERAMALAN TINGKAT KEDISIPLINAN MAHASISWA

Henderi[1]
Rizal Loa Wanda[2]

Email :*henderi@raharja.info , rizal.loa@raharja.info*

ABSTRAK

Data mining merupakan sebuah proses penggalian atau penemuan informasi baru dengan mencari pola tertentu dari sejumlah besar database yang telah tersedia. Teknik ini dapat membantu dalam pemanfaatan kembali data tersebut. Penelitian ini bermaksud melakukan penggalian informasi terhadap data absensi online di Perguruan Tinggi Raharja. Penggalian informasi dilakukan untuk meramalkan tingkat kedisiplinan mahasiswa. Penelitian ini bertujuan untuk mendapatkan set data yang berkualitas dan siap untuk dilakukan proses data mining untuk digunakan dalam proses peramalan tingkat kedisiplinan mahasiswa. Penelitian ini dilakukan dengan tahapan melakukan preprocessing data yang tediri dari pembersihan, ekstraksi, transformasi data, dan pemilihan atribut. Preprocessing dilakukan terhadap set data absensi online semester genap 2014/2015 dan set data induk mahasiswa. Hasil preprocessing mendapatkan, 1.836 record mahasiswa. Setiap record berisi variabel y sebagai atribut target yaitu kedisiplinan mahasiswa, dan 8 variabel x sebagai atribut prediktor yang diasumsikan memiliki pengaruh terhadap variabel y.

Kata Kunci: *Data mining, preprocessing, peramalan, tingkat kedisiplinan mahasiswa*

PENDAHULUAN

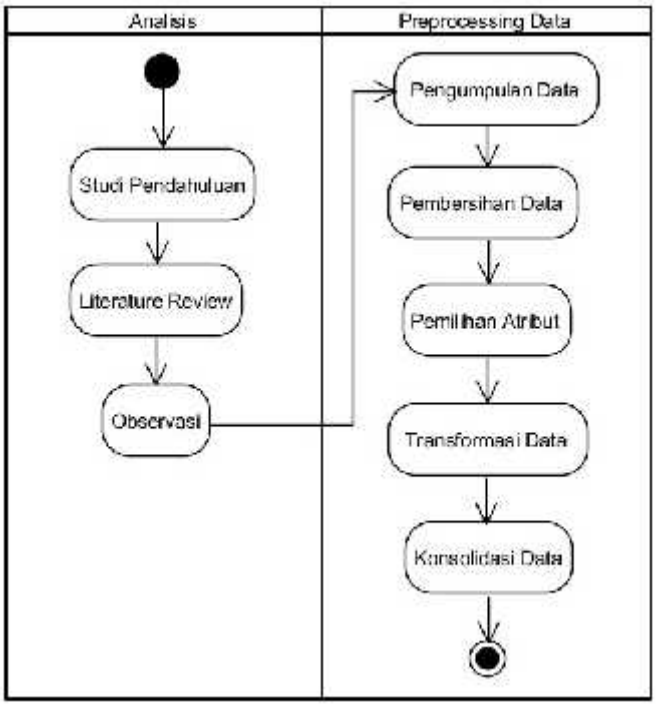
Perkembangan informasi di era digital ini terus meningkat setiap waktunya. Dampak perkembangan informasi tersebut ikut memberikan pengaruh yang besar terhadap peningkatan volume data yang disimpan dalam database. Data yang tersimpan setiap harinya menyebabkan penumpukan data yang besar. Jika hal tersebut tidak dimanfaatkan, maka data tersebut hanya akan menjadi data yang tidak terpakai di media penyimpanan. Tan mendifinisikan data miningsebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar [1]. Dengan *data mining*, penumpukan data dalam *volume* besar dapat dimanfaatkan kembali sebagai penunjang keputusan ataupun untuk keperluan prediksi di masa depan. Namun pada nyatanya, data dalam *volume* besar adalah hal yang kotor. [2] *Preprocessing*diperlukan karena data dunia nyata umumnya tidak lengkap (kurang nilai atribut, kurang atribut tertentu yang menarik, atau hanya berisi data agregat), kotor / *noise* (mengandung kesalahan atau *outlier*), dan tidak konsisten (mengandung perbedaan dalam kode atau nama).Hal ini memiliki pengaruh terhadap hasil dari proses data mining. Kualitas data yang digunakan untuk proses *data mining*, berbanding lurus dengan hasil yang akan didapat.

Data Absensi Online (AO) di Perguruan Tinggi Raharja, adalah data yang cukup potensial untuk digali pengetahuannya. Data yang terdapat di AO dapat digunakan kembali sebagai salah satu sumber data dalam proses data mining untuk meramalkan tingkat kedisiplinan mahasiswa. Mengingat data dalam dunia nyata adalah hal yang

kotor begitu juga dengan data AO, maka diperlukannya sebuah proses yang dapat meningkatkan kualitas dari data yang akan diolah. Salah satu proses yang ada dalam *Knowledge Discovery in Database*(KDD) dapat mengatasi masalah tersebut, proses itu adalah *Preprocessing* data.

METODOLOGI PENELITIAN

Preprocessing data untuk sistem peramalan tingkat kedisiplinan mahasiswa dibagi menjadi dua tahap aktivitasutama. Aktivitas pertama yaitu analisis yang terdiri dari studi pendahuluan, literature review, dan observasi. Observasi dilakukan di Perguruan Tinggi Raharja pada bagian RME yang terkait dengan objek penelitian yaitu data absensi online. Sedangkan aktivitas kedua adalah *preprocessing* data yang terdiri dari pengumpulan sumber data, pembersihan data, pemilihan atribut,transformasi data dan konsolidasi data yaitu replikasi data dari lebih dari set data untuk menjadi set data yang baru.



Gambar 1. Diagram Alir Metodologi Penelitian

LITERATURE REVIEW

Terdapat beberapa penelitian terkait mengenai penelitian ini. Diantaranya yang dilakukan oleh Astuti, dkk [3] mengenai prediksi siswa yang berpotensi melakukan ketidakdisiplinan menggunakanAlgoritma *Naïve Bayes Classifier*. [3] Penelitiannya menggunakan data asli siswa yang berjumlah 920 *record* sebagai sumber data dan atribut yang dipilih berjumlah 16 diantaranya 15 merupakan atribut prediktor dan 1 atribut target (ketidakdisiplinan).Sedangkan Mustafidah[4] menggunakan model Regresi untuk mengetahui hubungan antara motivasi belajar dengan tingkat kedisiplinan

agar dapat diprediksi. Sumber data dan atribut yang digunakan pada penelitiannya[4] berasal dari angket yang disebar pada 127 mahasiswa dan atribut yang digunakan hanya 2 yaitu motivasi belajar (variabel x) dan tingkat kedisiplinan (variabel y).

Penelitian terkait berikutnya yaitu penelitian Elisa [5] mengenai faktor-faktor yang mempengaruhi kedisiplinan. Objek penelitiannya [5] adalah kedisiplinan kerja karyawan yang bekerja di PT Suka Fajar Pekanbaru dengan sumber data berasal dari wawancara dan kuesioner dengan sampel sebanyak 74 orang responden menggunakan teknik cluster sampling. Penelitiannya [5] menjelaskan bahwa variabel bebas (*independent variable*) adalah faktor kompensasi, sanksi hukum dan kepemimpinan, sedangkan variabel terikatnya (*dependent variable*) adalah kedisiplinan kerja karyawan. Tujuan penelitiannya [5] adalah untuk mengetahui faktor-faktor yang mempengaruhi kedisiplinan kerja karyawan, dan untuk mengetahui faktor dominan yang mempengaruhi kedisiplinan kerja. Namun pada penelitiannya belum melalui proses data mining, yang sebenarnya dapat dikembangkan kembali menggunakan data mining. Sebagai contoh untuk meramalkan kedisiplinan kerja, sehingga tidak terbatas hanya untuk mengetahui faktor yang mempengaruhi kedisiplinan saja.

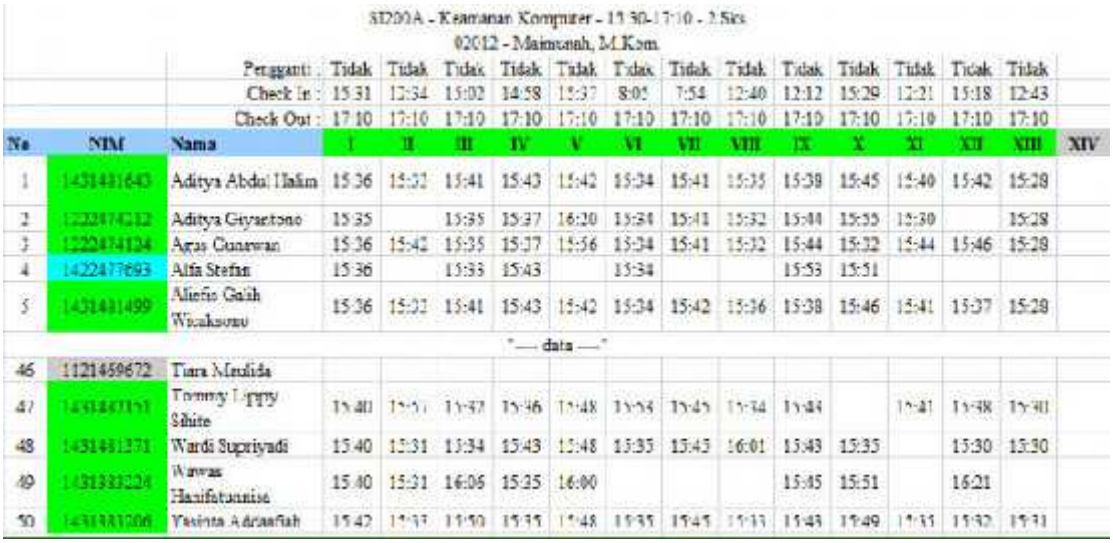
Penelitian berbeda telah dilakukan Ginting, dkk [6], telah membahas mengenai prediksi masa studi mahasiswa berdasarkan data nilai akademik. Penelitiannya menggunakan pendekatan yang ada di KDD, dan juga Algoritma C4.5 dalam penentuan masa studi mahasiswa, serta pohon keputusan (*decicion tree*) untuk melihat kemungkinan mahasiswa yang lulus lebih dari 8 semeseter. Serupa dengan penelitian sebelumnya[6], di penelitian Trayasiwi [7] juga membahas mengenai prediksi kelulusan mahasiswa dan menggunakan pendekatan yang ada di KDD. Namum perbedaannya terletak pada metode yang digunakannya yaitu menggunakan metode klusterisasi serta algoritma *K-Means*. [7] berpendapat bahwa data yang telah diklasterisasi tersebut menghasilkan kategori prediksi kelulusan mahasiswa berdasarkan lama atau tidaknya waktu kelulusan dan tinggi rendah IPK yang diperoleh mahasiswa pada setiap klaster.

PEMBAHASAN

A. Sumber Data

Pada penelitian ini, sumber data yang digunakan berasal dari dua sumber yang berbeda. Sumber pertama yaitu data Absensi Online (selanjutnya disebut AO) yang sudah tidak digunakan kembali (data historis). Pada studi kasus ini, data AO yang digunakan yaitu data pada semester genap tahun akademik 2014/2015 dengan hanya menggambil perkuliahan kelas teori yang memiliki bobot 2 dan 3 sks. Sumber data AO ini terdiri dari 338 kelas teori dengan tenaga pengajar sebanyak 119 dosen. Dan setiap data kelas tersebut (Gambar 2) berisi *record* waktu kehadiran mahasiswa disetiap pertemuan dengan jumlah maksimal 14 kali pertemuan disetiap semesternya.

Sumber data kedua yaitu data mahasiswa yang terdapat di SiS+ (*Students iLearning Services*) Perguruan Tinggi Raharja. Sumber data kedua ini diambil dari *web service* yang telah *on-line* (Gambar 3), sehingga data tersebut dapat diambil dimana saja dan kapan saja selama terhubung dengan internet.



SI209A - Keamanan Komputer - 13.30-17.10 - 2 Sks
02012 - Maimunah, M.Kom.

		Pengganti :	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
		Check In :	15:31	12:34	15:02	14:58	15:37	8:05	7:54	12:40	12:12	15:29	12:21	15:18	12:43
		Check Out :	17:10	17:10	17:10	17:10	17:10	17:10	17:10	17:10	17:10	17:10	17:10	17:10	17:10
No	NIM	Nama	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
1	1431481643	Aditya Abdul Halim	15:36	15:32	15:41	15:43	15:42	15:34	15:41	15:35	15:38	15:45	15:40	15:42	15:28	
2	1222474132	Aditya Giyantono	15:35		15:35	15:37	16:20	15:34	15:41	15:32	15:44	15:55	15:39		15:28	
3	1222474124	Agus Gunawan	15:36	15:42	15:35	15:37	15:56	15:34	15:41	15:32	15:44	15:32	15:44	15:46	15:28	
4	1422477693	Alfa Stefan	15:36		15:33	15:43		15:34			15:53	15:51				
5	1431481499	Alnefo Gaih Wisaksono	15:36	15:32	15:41	15:43	15:42	15:34	15:42	15:36	15:38	15:46	15:41	15:37	15:28	
							"— data —"									
46	1121459672	Tiara Maulida														
47	1431481151	Tommy Happy Sihite	15:40	15:31	15:37	15:46	15:48	15:34	15:43	15:34	15:44		15:41	15:38	15:30	
48	1431481371	Wardi Supriyadi	15:40	15:31	15:34	15:43	15:48	15:35	15:43	16:01	15:43	15:35		15:30	15:30	
49	1431333224	Wawas Hanifatunnisa	15:40	15:31	16:06	15:25	16:00				15:35	15:51		16:21		
50	1431331006	Yasinta Adriantiah	15:42	15:37	15:50	15:35	15:48	15:35	15:45	15:33	15:43	15:49	15:33	15:32	15:31	

Gambar 2. Sumber Data AO

Daftar WEB SERVICE

Show 10 entries

Search:

No	Nama Tabel
1	tbl_IMDayaKuliah
2	tbl_IMDosen
3	tbl_IMForm
4	tbl_IMGolonganDosen
5	tbl_IMGrading
6	tbl_IMIlan
7	tbl_IMJabatan
8	tbl_IMJadwalKJRS
9	tbl_IMJenisKelas
10	tbl_IMJenisRuang

Struktur tbl_TMMahasiswa

Show . Count. | Data / Jso |

No	Nama Field	Type	Panjang
1	NIM	string	10
2	NIPR	string	12
3	NamaDepan	string	30
4	NamaBelakang	string	70
5	NamaPanggilan	string	25
6	TempatLahir	string	50
7	TanggalLahir	date	10
8	KTP	string	16
9	KodeNegara	string	2
10	JenisKelamin	string	1
11	StatusPerkawinan	int	1

Gambar 3. Sumber Data Mahasiswa dari Web Service Raharja

B. Pembersihan Data

Pembersihan data pada studi kasus ini diantaranya adalah pembersihan data kelas dan pembersihan *record* mahasiswa. Pembersihan kelas dilakukan jika terdapat kelas yang diduga abnormal maka kelas tersebut dihapus dan pembersihan *record* mahasiswa jika mahasiswa tersebut tidak hadir satu kali pun selama 14 pertemuan perkuliahan.

1. Pembersihan Kelas

Ada 338 kelas teori yang digunakan dalam proses pembersihan, terdapat 4 kelas yang dikatergorikan sebagai data anomali data. Diindetifikasi sebagai data anomlai karena absensi penuh disetiap pertemuan dan waktu absensi ada di waktu yang sama.

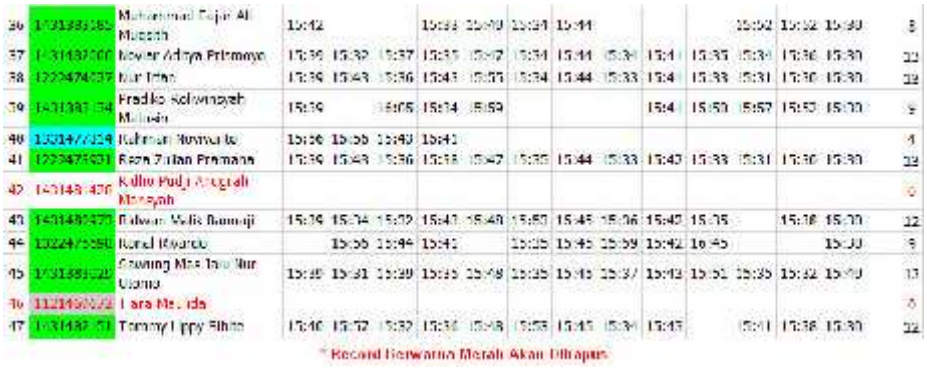
		Pertemuan	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
		Check In	8.00	7.50	7.18	7.32	8.17	7.55	7.55	7.39	7.40	7.42	8.00	7.49	7.51	8.00
		Check Out	9.40	9.40	9.40	9.40	9.40	9.40	9.40	9.40	9.40	9.40	9.40	9.40	9.40	9.40
No	NIM	Nama	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
1	1312483749	Anis Khoiriyah Nisa	8.12	8.08	8.00	8.00	8.00	8.16	8.14	7.49	7.41	8.11	8.00	8.25	8.24	8.00
2	1412482024	Ayu Intan Marsyati	8.12	8.08	8.00	8.00	8.00	8.16	8.14	7.49	7.41	8.11	8.00	8.25	8.24	8.00
3	1412483121	Chaerunisa Syviri	8.12	8.08	8.00	8.00	8.00	8.16	8.14	7.49	7.41	8.11	8.00	8.25	8.24	8.00
4	1312476375	Chelsie Indrie Pramesivi	8.12	8.08	8.00	8.00	8.00	8.16	8.14	7.49	7.41	8.11	8.00	8.25	8.24	8.00
5	1312481086	Dery Amilia	8.12	8.08	8.00	8.00	8.00	8.16	8.14	7.49	7.41	8.11	8.00	8.25	8.24	8.00
6	1312483014	Devi Septiani	8.12	8.08	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.25	8.24	8.00
7	1312477393	Dian Triyanjaya	8.12	8.08	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.25	8.24	8.00
8	1312483673	Dimas Singgih Prasetya	8.12	8.08	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.25	8.24	8.00
9	1412483208	Fachri Ardiansyah	8.12	8.08	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.25	8.24	8.00
10	1412482145	Faisal Haris Ridwan	8.12	8.08	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.25	8.24	8.00
11	1412483357	Fitria Supriyanasih	8.12	8.08	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.17	8.14	8.00
12	1312481750	Fitri Yunita Anita	8.12	8.08	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.17	8.14	8.00
13	1312483051	Hanika Yuniasa	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.17	8.14	8.00
14	1312476831	Hartono Widjaja	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.17	8.14	8.00
15	1312482479	Jessica Kaisarsari	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.17	8.14	8.00
16	1412482514	Mamay Hamamah	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.17	8.14	8.00
17	1412482859	Mila Tryulani Karina	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.17	8.14	8.00
18	1312483502	Muharrad Arif Nurdin	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.17	8.14	8.00
19	1412478655	Nur Farida	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.17	8.14	8.00
20	1412482699	Nur Hasanah	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.50	7.41	8.11	8.00	8.17	8.14	8.00
21	1312477956	Nur Rahmat Aditya Pratama	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.51	7.41	8.11	8.00	8.17	8.14	8.00
22	1312482230	Nurlaila	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.51	7.41	8.11	8.00	8.17	8.14	8.00
23	1312483532	Pandi Budiana	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.51	7.41	8.11	8.00	8.17	8.14	8.00
24	1412481985	Prayudha Bima Hernyanto	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.51	7.41	8.11	8.00	8.17	8.14	8.00
25	1312477027	Rosa Wulandari	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.51	7.41	8.11	8.00	8.17	8.14	8.00
26	1412482921	Rinka Permata Dewi	8.12	8.09	8.00	8.00	8.00	8.16	8.14	7.51	7.41	8.11	8.00	8.17	8.14	8.00
27	1412481017	Silva Ayu Lestari	8.12	8.09	8.00	8.00	8.00	8.16	8.15	7.51	7.41	8.11	8.00	8.17	8.15	8.00
28	1412481825	Ujat Sujana	8.12	8.09	8.00	8.00	8.00	8.16	8.15	7.51	7.41	8.11	8.00	8.17	8.15	8.00
29	1312483011	Yanti Nurmalasari	8.12	8.09	8.00	8.00	8.00	8.16	8.15	7.51	7.41	8.11	8.00	8.17	8.15	8.00

Gambar 4. Contoh Kelas yang Abnormal

Gambar 4 adalah contoh data anomaly karena absensi terisi penuh selama 14 kali pertemuan dengan waktu absen bersamaan. Data ini diragukan kebenaran absensinya, dan jika data kelas tersebut digunakan dapat mengakibatkan ketidak-akuratan pada hasil dari proses data mining.

2. Perhitungan Variabel Y dan Pembersihan Record Mahasiswa

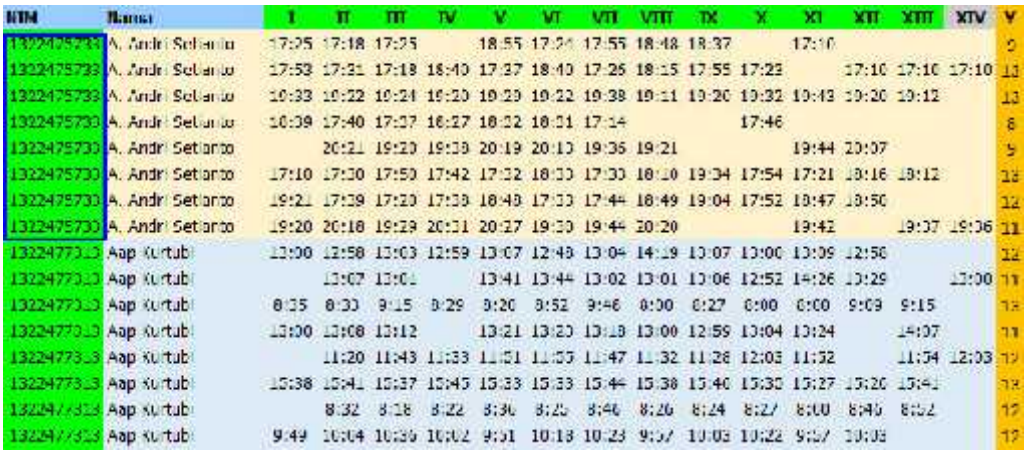
Perhitungan atau *counting* variabel y (kehadiran mahasiswa) diperlukan untuk mengetahui jumlah kehadiran mahasiswa dikelas. Selain untuk pembersihan *record* mahasiswa, perhitungan variabel y juga akan digunakan untuk menghasilkan rata-rata kehadiran yang akan disebut kedisiplinan mahasiswa. Pada studi kasus ini telah ditentukan untuk pembersihan *record* mahasiswa adalah jika pada *record* mahasiswa dikelas tersebut 0 (nol) kehadiran (Gambar 5) selama 14 kali pertemuan, maka *record* tersebut dihapus dan tidak digunakan.



Gambar 5. Contoh Record Mahasiswa yang 0 Kehadiran

3. Penggabungan Record Untuk Mencari Rata-Rata Variabel Y

Satu mahasiswa dapat mengambil banyak mata kuliah (kelas) disetiap semester, sehingga hal yang perlu dilakukan adalah penggabungan (*Grouping*) dari setiap kehadiran mahasiswa (variabel y) di kelas-kelas mata kuliah sesuai dengan NIM dan nama mahasiswa.



Gambar 6. Sebelum Penggabungan Record

Dari Gambar 6 dapat dilihat bahwa, disetiap *record* adalah *record* mahasiswa yang terdapat di 1 kelas mata kuliah. Sebagai contoh mahasiswa dengan nim 1322475733 memiliki 8 variabel y, dengan kata lain mahasiswa tersebut pada semester ini mengambil 8 mata kuliah. Sehingga setelah proses *grouping* maka hasilnya adalah setiap mahasiswa memiliki nilai rata-rata kehadiran / rata-rata variabel y ($\bar{y}$).

NIM	NAMA	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	y11	ȳ
1322475733	A. Andri Setianto	9	13	13	8	9	13	12	11				11.00
1322477313	Aap Kurtubi	12	11	13	11	12	13	12	12				12.00
1322476305	Abdul Azis	10	10	11	13	9	10	12					10.71
1414481555	Abdul Azis	14	14	14	14	12	12	12	12	14			12.78
1411481283	Abdul Azis	11	10	10	10	11	9	13	13	12			11.00
1133488559	Abdul Fatah	10	11	11	13	10							11.00
1322477015	Abdul Haqy Aji Prastian	12	11	13	13	13	11	12	13	14			13.11
1222474392	Abdul Mujid	12	10	13	12	12	11	11	12				11.63
1321476870	Abdul Mukarobi	12	8	10	8	13	6	7					9.14
1433481330	Abdul Rizki Zaikasy	14	13	13	13	13	13	12					13.00
1011476941	Abdul Rosid	13	9	12	8	7							9.80
1433481636	Abdul Waci Rahmat	13	12	13	12	14	14	14					13.14
1212473737	Abdurrahman Muhammad Ghozali	10	9	8	9	9	10	9	8				9.00
1531483314	Abdurrauf Rafli	13	11	9	9	12	9	11	8				10.25

Gambar 7. Hasil Dari Penggambungan dan Nilai ȳ

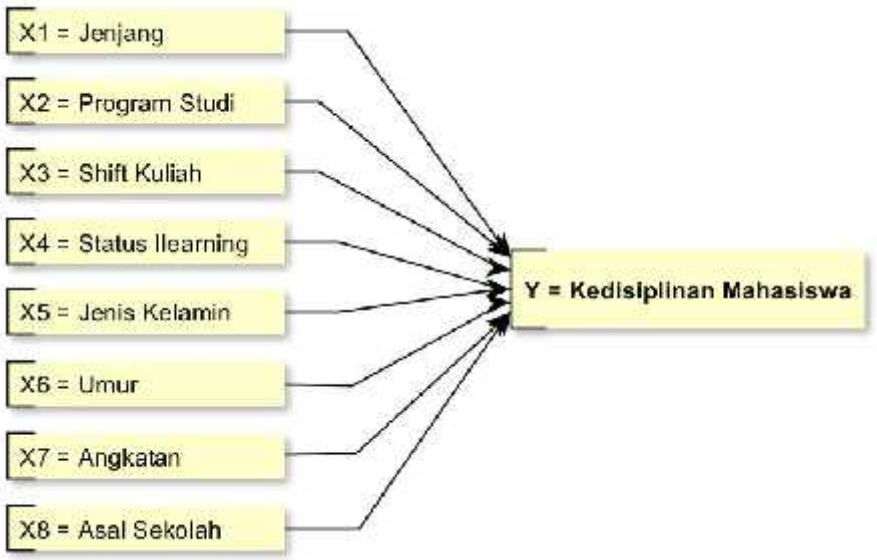
C. Pemilihan Atribut

Pemilihan atribut atau variabel x harus sesuai dengan pertimbangan bahwa atribut tersebut terdapat pengaruh dan relevan terhadap variabel y. Dalam penelitian ini, telah dipilih beberapa atribut yang diasumsikan berpengaruh terhadap tingkat kedisiplinan mahasiswa (Tabel 1), diantaranya Jejang kuliah, program studi, shift kuliah, status pekerjaan, status ilearning, jenis kelamin, umur, angkatan, dan asal sekolah. Namun saat penarikan data untuk atribut status pekerjaan, data yang didapat kurang mencukupi, sehingga atribut tersebut tidak dapat digunakan.

Tabel 1. Variabel X

X	Atribut
X1	Jenjang Kuliah
X2	Program Studi
X3	Shift Kuliah
X4	Status iLearning
X5	Jenis Kelamin
X6	Umur
X7	Angkatan
X8	Asal Sekolah

Dan jika dilihat dari hubungannya (Gambar 8), variabel y dengan variabel x memiliki hubungan asimetris lebih dari dua variabel atau dapat disebut hubungan variabel *multivariat.*



Gambar 8. Hubungan Antar Variabel Multivariat

D. Penarikan Data Untuk Variabel X

Penarikan data untuk variabel x bersumber data yang berbeda dengan set data AO. Sumber data tersebut didapat dari *web service* yang disediakan oleh Perguruan Tinggi Raharja untuk keperluan riset mahasiswa. Format pertukaran data yang disediakan berupa JSON (*JavaScript Object Notation*) sehingga perlu dilakukan ekstraksi data.

Ekstraksi Data

Dikarenakan format data yang digunakan berbeda dari sumber data maka harus ada perubahan format serta dilakukan penyimpanan sementara hasil ekstraksi untuk penggabungan dari sumber data lain (beda tabel).

Gambar 9. Contoh JSON Dari Tabel-Tabel Mahasiswa

Gambar 9 adalah contoh dari format JSON yang dihasilkan dari web service yang datanya dapat diolah. Sebagai contoh, berikut adalah bagaimana cara mengambil data dari web service tersebut.

Contoh : pada tabel tbl_TMMahasiswa, data diambil dari JSON pada url

http://rapi.raharja.me/JSON/qTWfK1EAGJSbLKAcp3qu/nim/NIM_MAHASISWA

Di ekstraksi menggunakan program sederhana dari bahasa pemrograman PHP.Berikut adalah *listing program*-nya :

```
<?php
$nim = "1322476037";
$url = "http://rapi.raharja.me/JSON/qTWfK1EAGJSbLKAcp3qu/nim/".$nim;
$file = file_get_contents($url);
$data = json_decode($file);
echo "<table border=1>
              <tr>
                     <th>Nama</th>
                     <th>Jenis Kelamin</th>
                     <th>Tanggal Lahir</th>
              </tr>
              <tr>";
foreach($data as $key) {
       echo '<td>'.$key->NamaDepan.' '.$key->NamaBelakang.'</td>';
       echo '<td>'.$key->JenisKelamin.'</td>';
       echo '<td>'.$key->TanggalLahir.'</td>';
}
echo "</tr>
       </table>";
?>
```

Maka output dari *listing program* tersebut adalah sebagai berikut :

Tabel 2. Contoh Hasil Output Dari Penarikan Variabel X

Nama	Jenis Kelamin	Tanggal Lahir
Rizal Loa Wanda	L	1994-12-20

Kelebihan dari format pertukaran data JSON adalah data tersebut dapat disesuaikan dengan format yang diinginkan, dan hampir semua bahasa pemrograman dapat merepresentasikan data dari format JSON ini (tidak terbatas oleh *platform*).

NIM	NAMA	Jenjang	Program Studi	Shift Kuliah	iLearning	Jenis Kelamin	Umur	Angkatan	Asal Sekolah
1322476616	Riyan Nova Saputra	100	120	1	0	L	11/27/1995	13	SMAN 4 TANGERANG
1422381588	Riyani Juniarti	300	330	1	0	P	6/24/1995	14	
1322476837	Rizal Loa Wanda	400	430	1	0	L	12/20/1994	13	SMK NEGERI 1 TANGERANG
1412482927	Rizka Permata Devi	400	410	1	1	P	12/21/1996	14	
1122468923	Rizka Septiandy	100	120	2	0	L	9/20/1990	11	SMK Penerbangan Padang
1022464650	Rizki Adigana	100	120	1	0	L	10/27/1991	10	SMA 1 Pasarkemis
1314476702	Rizki Afri Dani Firmansyah	400	410	1	0	P	4/20/1995	13	SMAN 11 TANGERANG
1431482286	Rizki Aprian	400	430	1	0	L	4/17/1994	14	
1433482687	Rizki Aulia Ramdhani	400	430	1	1	L	2/20/1997	14	
1322476834	Rizki Imam Wahyudi	100	120	1	0	L	12/26/1995	13	SMAN 11 KABUPATEN TANGERANG
1411482101	Rizki Mandala Anugerah	300	310	2	0	L	5/25/1990	14	
1214473314	Rizki Regina Ulfauzah	400	410	2	0	P	12/22/1995	13	SMAN 17 KABUPATEN TANGERANG
1314476043	Rizki Setyawan	400	410	2	0	L	3/5/1994	13	SMKN 2
1414482193	Rizki Widi Darmawan	300	310	1	1	L	6/29/1996	14	
1414483001	Riskiana Nurhidayah	100	110	2	0	P	2/5/1995	14	
1212472729	Rizky Agustian	100	110	1	1	L	8/17/1995	12	SMA Yuppentek 4
1333477047	Rizky Indra Kurniawan	400	430	1	1	L	8/28/1995	14	SMK Nusa Putra
1122467624	Rizky Martin	400	420	2	0	L	1/21/1994	11	SMAN 6 TANGERANG

Gambar 10. Hasil Ekstraksi Data Dari Web Service

Sehingga hasil ekstraksi dari beberapa tabel pada set data mahasiswa di *web service* menjadi seperti di Gambar 10, dimana dari format pertukaran data JSON menjadi format tabel yang mengisi setiap atribut x pada *record* mahasiswa.

E. Transformasi Data

Transformasi data dilakukan agar nilai dari atribut-atribut yang sebelumnya masih berupa kode atau id, menjadi data kategorikal yang dapat dipahami (Gambar 11).

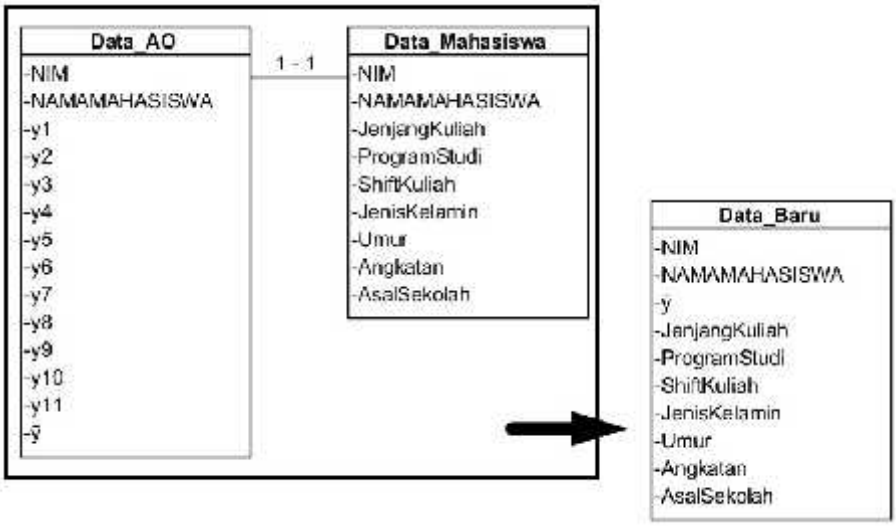
1. Pada atribut jenjang, dan program studi, perubahan data atribut dari kode menjadi data kategorikal didapat dari tabel yang terdapat pada *web service* Raharja.
2. Perubahan data pada atribut shift kuliah ialah dari data angka menjadi data kategorikal : 1 = siang, dan 2 = malam.
3. Perubahan data pada atribut iLearning ialah *discretization* dimana dari bentuk biner (0 dan 1) menjadi Tidak dan Ya.
4. Perubahan data pada atribut umur ialah dari yang sebelumnya tanggal lahir menjadi umur.

NIM	Nama	Jenjang	Program Studi	Shift Kuliah	iLearning	Jenis Kelamin	Umur	Angkatan	Asal Sekolah
1322476616	Riyan Nova Saputra	S1	Teknik Informatika	Siang	Tidak	L	20	13	SMAN 4 TANGERANG
1422381588	Riyani Juniarti	D3	Komputerisasi Akuntansi	Siang	Tidak	P	20	14	
1322476837	Rizal Loa Wanda	S1	Teknik Informatika	Siang	Tidak	L	21	13	SMK NEGERI 1 TANGERANG
1412482927	Rizka Permata Devi	S1	Sistem Informasi	Siang	Ya	P	19	14	
1122468923	Rizka Septiandy	S1	Teknik Informatika	Malam	Tidak	L	25	11	SMK Penerbangan Padang
1022464650	Rizki Adigana	S1	Teknik Informatika	Siang	Tidak	L	24	10	SMA 1 Pasarkemis
1314476702	Rizki Afri Dani Firmansyah	S1	Sistem Informasi	Siang	Tidak	P	21	13	SMAN 11 TANGERANG
1431482286	Rizki Aprian	S1	Sistem Komputer	Siang	Tidak	L	22	14	
1433482687	Rizki Aulia Ramdhani	S1	Sistem Komputer	Siang	Ya	L	19	14	
1322476834	Rizki Imam Wahyudi	S1	Teknik Informatika	Siang	Tidak	L	20	13	SMAN 11 KABUPATEN TANGERANG
1411482101	Rizki Mandala Anugerah	S1	Sistem Informasi	Malam	Tidak	L	25	14	
1214473314	Rizki Regina Ulfauzah	S1	Sistem Informasi	Malam	Tidak	P	22	12	SMAN 17 KABUPATEN TANGERANG
1314476043	Rizki Setyawan	S1	Sistem Informasi	Malam	Tidak	L	21	13	SMKN 2
1414482193	Rizki Widi Darmawan	D3	Manajemen Informatika	Siang	Ya	L	19	14	
1414483001	Riskiana Nurhidayah	S1	Sistem Informasi	Malam	Tidak	P	21	14	
1212472729	Rizky Agustian	S1	Sistem Informasi	Siang	Ya	L	21	12	SMA Yuppentek 4
1333477047	Rizky Indra Kurniawan	S1	Sistem Komputer	Siang	Ya	L	22	13	SMK Nusa Putra
1122467624	Rizky Martin	S1	Teknik Informatika	Malam	Tidak	L	22	11	SMAN 6 TANGERANG

Gambar 11. Hasil Dari Transformasi Data

F. Konsolidasi Data

Konsolidasi data adalah replikasi data lebih dari satu set data untuk menjadi satu set data. Pada studi kasus ini, set data *absensi online* yang telah diketahui rata-rata dari variabel y yaitu kedisiplinan mahasiswa, dan set data mahasiswa yang berisi atribut-atribut atau variabel x akan digabung menjadi satu set data baru.



Gambar 12. Skema dari Konsolidasi Data

Gambar 12 adalah hubungan antara data_*absnsi online* dengan data_mahasiswa yang telah diolah dan digabung menjadi satu. Penggabungan ini menghasilkan set data baru yang siap untuk diolah dalam data mining (Gambar 13).

NIM	Nama	Rata Rata V	Jenjang	Program Studi	Shift Kuliah	Status iLearning	Jenis Kelamin	Umur	Angkatan	Asal Sekolah
13224757[illegible]	A. Andri Setiawan	11.00	S1	Teknik Informatika	Malam	Tidak	L	28	16	SMAN 1 Cisoka
13[illegible]7813	Aap Kurniati	12.00	S1	Teknik Informatika	Siang	Tidak	P	27	16	SMAN 19 KABUPATEN TANGERANG
1411481283	Abdul Aziz	11.00	S1	Sistem Informasi	Malam	Tidak	L	21	14	
1411481555	Abdul Aziz	12.78	S1	Sistem Informasi	Malam	Tidak	L	21	14	
1322476305	Abdul Aziz	10.71	S1	Teknik Informatika	Malam	Tidak	L	26	13	
1133468559	Abdul Fatah	11.00	S1	Sistem Komputer	Siang	Ya	L	21	11	MA Darus Sadah
13[illegible]7045	Abdul Hagy Ajl Pramana	13.11	S1	Teknik Informatika	Siang	Tidak	L	30	16	SMK Awaloh
1322474392	Abdul Mujib	11.88	S1	Teknik Informatika	Siang	Tidak	L	23	13	SMK PGRI 1 Balaraja
1321476870	Abdul Mukaroki	9.14	S1	Teknik Informatika	Siang	Ya	L	21	13	SMK PUSTEK SERPONG
1433481390	Abdul Rizki Zarkasy	13.00	S1	Sistem Komputer	Siang	Ya	L	20	14	
13114[illegible]941	Abdul Rosid	9.00	S1	Sistem Informasi	Malam	Tidak	L	25	13	
1433481686	Abdul Waqi Rahmat	13.14	S1	Sistem Komputer	Siang	Ya	L	23	14	
1212473737	Abdurrahman Muhammad Ch	9.00	S1	Sistem Informasi	Siang	Ya	L	21	12	SMA Yuppentek 1
1531403314	Abdurrauf Rafli	10.25	S1	Sistem Komputer	Malam	Tidak	L	26	15	
1422481678	Abdus Samad	13.44	S1	Teknik Informatika	Siang	Tidak	L	19	14	
14[illegible]8806	Abie Marchelo Mangol	12.22	S1	Teknik Informatika	Siang	Tidak	L	20	14	
1321473800	Abigail Arni	11.57	S1	Teknik Informatika	Siang	Ya	L	21	13	SMAN 13 KABUPATEN TANGERANG
1133468076	Achmad Bahaeni A	11.75	S1	Sistem Komputer	Siang	Ya	L	23	11	SMAN 15 Tangerang
13s13[illegible]6143	Achmad Dardabi	10.78	D3	Teknik Informatika	Malam	Tidak	L	25	13	SMK YP Karya 2
12214[illegible]3100	Achmad Khairul Fadli	13.00	S1	Teknik Informatika	Malam	Tidak	L	22	12	SMKN 1 Kab.Tangerang
1321473805	Achmad Kurniawan	9.71	S1	Teknik Informatika	Siang	Ya	L	20	13	SMA ISLAMIC CENTRE
1422482719	Achmad Rifki Arsyadi	11.80	S1	Teknik Informatika	Siang	Tidak	L	20	14	
1322475440	Achmad Robby Farsa	9.63	S1	Teknik Informatika	Malam	Tidak	L	21	13	SMAN 17 KABUPATEN TANGERANG
12s13[illegible]785	Achmad Sobirawan	10.50	D3	Teknik Informatika	Malam	Tidak	L	26	12	SMKN 6 Serpong
12214[illegible]946	Achmad Subekti	9.75	S1	Teknik Informatika	Siang	Ya	L	21	12	SMA Nusantara 1
1422482214	Achmad Wais Alqorni	11.50	S1	Teknik Informatika	Malam	Tidak	L	20	14	
1322476604	Achmadsyah Ilanov Dwi Prabu	11.50	S1	Teknik Informatika	Malam	Tidak	L	22	13	SMKN 6 Kabupaten Tangerang
12214[illegible]903	Adco Priyandi	11.89	S1	Teknik Informatika	Siang	Ya	L	22	12	SMK PGRI 109 Tangerang
13[illegible]6870	Adeh Setiyoni	11.40	S1	Teknik Informatika	Siang	Ya	P	20	13	SMKN 4 TANGERANG
1121466961	Adam Maulana Rachman	9.25	S1	Teknik Informatika	Siang	Tidak	L	24	11	SMK Vochtech 1
1321476078	Adang Arief Runadi	11.40	S1	Teknik Informatika	Siang	Ya	L	21	13	PGMI Bina Warga
1331476614	Ade Ferdianayah	13.38	D3	Teknik Informatika	Malam	Tidak	L	25	13	SMK Komputer
13113[illegible]249	Ade Hikmah Suryani	9.50	D3	Manajemen Informatika	Malam	Tidak	P	20	13	SMK Al Rahman
12214[illegible]925	Ade Kosasih	13.00	S1	Teknik Informatika	Siang	Ya	L	21	12	SMK AL HALIMAH
1121469073	Ade Maulana Khumaedi	9.75	S1	Teknik Informatika	Siang	Tidak	L	21	11	SMK N 3 PANGKALAN
1411480726	Ade Permana	12.20	S1	Sistem Informasi	Siang	Tidak	L	26	14	
1411482274	Ade Solawudin	13.22	S1	Sistem Informasi	Siang	Tidak	L	19	14	
1411483296	Ade Sopah Fauzah	12.89	S1	Sistem Informasi	Siang	Tidak	P	19	14	
1322473700	Ade Yahan Hayumi	10.88	S1	Teknik Informatika	Malam	Tidak	L	23	13	SMK Ormatif Al-Hikma
0921463781	Adhi Dumyati	8.86	S1	Teknik Informatika	Siang	Tidak	L	25	09	SMAN 6 Tangerang
1011464611	Adhitya Dwi Pencasakti	10.30	S1	Sistem Informasi	Siang	Tidak	L	24	10	SMA Islamic Centre Tangerang
0814461864	Adi Adriansah	8.00	S1	Sistem Informasi	Malam	Tidak	L	26	08	SMAN 1 Balaraja
13[illegible]8838	Adi Alfian	13.44	S1	Sistem Informasi	Malam	Tidak	L	22	13	
1211473434	Adi Gunawan	10.38	S1	Sistem Informasi	Siang	Tidak	L	23	12	SMKN 2 Kabupaten Tangerang
1321477140	Adi Praseliyo	13.43	S1	Teknik Informatika	Siang	Ya	L	20	13	SMKN 5 KAB.TANGERANG

Gambar13. Hasil Dari Konsolidasi Data

KESIMPULAN

Preprocessing data telah dilakukan sesuai dengan proses yang ada pada KDD, dengan menghasilkan set data yang baru hasil dari penggabungan dua sumber data yang berbeda yaitu data absensi online mahasiswa dan data mahasiswa. Proses yang dilakukan diantaranya penarikan data dari sumber data, pembersihan data, pemilihan atribut, transformasi data, dan konsolidasi data. *Preprocessing* data pada studi kasus ini menghasilkan 1.836 *record* mahasiswa yang siap untuk diproses dalam data mining. Penelitian ini juga telah menentukan atribut-atribut yang diasumsikan memiliki pengaruh terhadap kedisiplinan mahasiswa (variabel y), diantaranya ialah jenjang kuliah, program studi, shift kuliah, status iLearning, jenis kelamin, umur, angkatan dan asal sekolah.

DAFTAR PUSTAKA

[1] D. Delen, *Real-World Data Mining: Applied Business Analytics and Decision Making*. New Jersey: Pearson Education, Inc, 2014.
[2] E. Prasetyo, DATA MINING - *Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: C.V ANDI OFFSET, 2012.
[3] G. Trayasiwi, "Penerapan Metode Klastering dengan Algoritma k-Means Untuk Prediksi Kelulusan Mahasiswa Pada Program Studi Teknik Informatika Strata Satu", Skripsi, Universitas Dian Nuswantoro, 2015.
[4] H. Mustafidah, "Model Regresi Data Mining Motivasi Belajar Pengaruhnya Terhadap Tingkat Kedisiplinan Mahasiswa", *JUITA*, vol. 1, no. 1, 2010.
[5] N. Astuti, K. Kurniasi and M. Arief, "Analisis Prediksi Tingkat Ketidakdiplinan Siswa Menggunakan Algoritma Naïve Bayes Classifier (Studi Kasus : SMK Negeri 1 Pacitan)", *Seminar Nasional Teknologi Informasi dan Multimedia*, vol. 3, no. 1, 2015.
[6] P. Elisa, "Analisis Faktor-Faktor Yang Mempengaruhi Kedisiplinan Kerja Karyawan Pada PT. Suka Fajar Pekanbaru", Skripsi, Universitas Islam Negeri Sultan Syarif Kasim Riau, 2013.
[7] S. Ginting, W. Zarman and I. Hamidah, "Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik", *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, 2014.